

Trabajo Fin de Grado

Modelos econométricos con datos de corte transversal: El precio del alquiler de la vivienda en Zaragoza

Autor

Helena Ibarra Gil

Director

Inmaculada Villanúa

Facultad de Economía y Empresa
2020

En primer lugar, me gustaría agradecer la ayuda que me ha prestado durante todo el trabajo a Inmaculada Villanúa, por haberme orientado y animado en todo momento, además de enseñarme a ser más metódica no sólo en la realización de este trabajo. Ha conseguido que el proceso fuera más ameno y constructivo.

También me gustaría agradecer en concreto a dos profesores, el papel que ha supuesto en mi formación. En primer lugar, a Gloria Jarne por mantener mi entusiasmo por las matemáticas en el primer año de carrera. Y, en segundo lugar, a Francisco Fatás por potenciarlo en el último año de carrera y por su gran vocación hacia los estudiantes. El manual de econometría que me regaló me ha ayudado a reforzar y ampliar lo aprendido en la carrera y a la realización de este trabajo.

También quiero mostrar mi agradecimiento a mi familia y amigos por escucharme, sobre todo a Lucía Jaqués que ha sabido aconsejarme y animarme en los peores momentos.

Por último, me gustaría darle las gracias al amor de mi vida, Sergio Martín Segura por su esfuerzo, paciencia y apoyo incondicional.

Abstract

En España, sólo el 18,5% de los jóvenes menores de treinta años consiguieron independizarse durante 2019. Las razones pueden encontrarse en la precariedad del mercado laboral, los obstáculos que encuentran al adquirir una vivienda o los precios de los alquileres. Respecto a estos últimos realizamos un estudio con el objetivo de conocer y cuantificar el efecto que determinados factores provocan sobre el precio de alquiler de pisos en Zaragoza. Utilizamos una base de datos de corte transversal para estimar la relación entre el precio y una serie de factores explicativos, a través de la metodología econométrica.

Los resultados que hemos obtenido corroboran las sospechas de partida. Los alquileres son más caros cuando: la zona es más rica, la vivienda tiene más superficie, el edificio tiene ascensor, hay más de un baño y se encuentran situados cerca del centro o de la universidad. Los dos factores con un mayor efecto son: la presencia de ascensor o de más de un baño.

Para concluir el trabajo, realizamos el análisis estructural con el modelo que, cumpliendo los requisitos para ser considerado adecuado, mejor explique los precios. Saber la magnitud y signo de determinados factores puede ayudar a tomar decisiones ya sea a nivel individual como institucional. También realizamos varias predicciones extramuestrales para unas determinadas características de la vivienda.

In Spain, only 18,5% of young people lower than thirty years old got to liberate in 2019. The reasons could be found into laboral market's insecurity, blockage for buying a housing and the rent prices. In relation to the last one, we made a project with the goal of knowing and quantifying the effects of certains factors that affect the price of rent in Zaragoza. We used a cross-section database to estimate the relationship of prices between explanatories factors through econometric methodology.

Our results confirm the starting suspicion. The rent are more expensive when: zone is richer, the housing is more spacious, building has elevator, there are more than one bathroom and housing is located close to the center or the university. The factors with more influence are: if it has elevator and more than one bathroom.

As a conclusion, we make the structural analysis with the model which better explains prices, while achieving the requirements for considering it suitable. Knowing measure and sign of certain factors can be helpful for decision making of individuals or institutions. We also make a few extra-sample prediction to certain characteristic of housing.

Índice

1. Introducción
2. Enfoque teórico
3. Metodología
4. Variables y datos
5. Modelos estimados
 - a. Modificaciones y mejoras
 - b. Modelos que superan la validación
6. Selección del modelo
7. Explotación
8. Conclusiones
9. Bibliografía

Introducción

Según el informe de López Oller (2019) desde 2011, los jóvenes se han encontrado con más obstáculos a la hora de emanciparse. Este autor, explica que en España “sólo el 18,5% de las personas entre 16 y 29 años están emancipados desde el punto de vista residencial”. También matiza que los movimientos migratorios pueden influir en las cifras de población emancipada.

Este bajo porcentaje puede deberse a múltiples razones, pero entre ellas destaca la situación laboral de los jóvenes, los estrictos requisitos que exigen las entidades bancarias para conceder créditos hipotecarios y los precios de los alquileres. Por este motivo, hemos elegido esta última como variable de estudio.

El mercado de trabajo juvenil se caracteriza por una mayor temporalidad y tasa de desempleo. Este autor afirma que, aunque los jóvenes tuvieran una situación laboral estable, el alquiler de una vivienda en España supondría el 94,4% de sus ingresos en 2019. Por ello, la solución que encuentran son los pisos compartidos, de forma que se dividan el gasto del alquiler de una misma vivienda. Aun así, estos jóvenes tendrían que dedicar entre el 30 y el 40% de su salario (siendo este porcentaje más alto para ciudades como Barcelona).

Las recomendaciones que López Oller (2019) plantea para aumentar las oportunidades de emancipación de los jóvenes son referidas principalmente a los precios de los alquileres y su calidad.

El principal objetivo de este trabajo es cuantificar la repercusión de diferentes factores, tanto externos como internos, sobre el precio de los alquileres en núcleos urbanos. Para ello planteamos un modelo econométrico con datos de corte transversal que corresponden a pisos ofertados de alquiler de la ciudad de Zaragoza. Hemos escogido dicha ciudad por su valor representativo donde múltiples marcas prueban sus productos, según M. F. ZARAGOZA (2014). Además, es importante destacar que en Aragón tenemos una proporción de jóvenes emancipados ligeramente superior a la media nacional, como comenta López Oller (2019). Ello puede deberse, según este autor a que más de la mitad

de los jóvenes asalariados tienen un contrato indefinido. Por tanto, en Aragón encuentran menos dificultades a la hora de acceder a un alquiler o una hipoteca.

Por otro lado, el momento temporal de nuestra base de datos corresponde a octubre de 2019. En este año, según Viaña (2019) la economía española, junto con la mundial, daba señales de desacelerarse. Para el caso español, las tensiones comerciales globales y la incertidumbre de la ejecución del Brexit pueden ser factores que afectan negativamente a su economía, de acuerdo con el informe del Banco de España (2019). No obstante, en dicho informe también se explica que, a pesar del lento crecimiento global, España mantiene un ritmo superior en relación con sus socios europeos.

La realización del modelo ha supuesto una oportunidad única para profundizar en el desarrollo de modelos econométricos y, de esta forma, valorar el esfuerzo que requieren la selección de la variable apropiada, el trabajo de recolección de datos, y la elaboración y selección de modelos. También me ha servido para interiorizar el modo de trabajo disciplinado y riguroso que demanda un proyecto de estas características.

En este contexto espacio-temporal se desarrolla el trabajo, que se estructura de la forma siguiente. En las secciones 2 y 3 se establece el enfoque teórico en el que se plantea el modelo, y la metodología que hemos seguido para llevarlo a cabo, respectivamente. La sección 4 se dedica a describir las variables y los datos que usaremos para en la sección 5 plantear diversas especificaciones y estimaciones. Posteriormente, la selección del mejor modelo, según el objetivo del investigador, se plantea en la sección 6. Con el mejor modelo se lleva a cabo la explotación, desde dos perspectivas diferentes, análisis estructural y predicción. La sección 8 resume los principales resultados del trabajo.

Enfoque teórico

Como ya hemos comentado, el objetivo de este trabajo es analizar qué variables explican el precio del alquiler de vivienda en Zaragoza por lo que para ello es necesario conocer qué hace que oscilen. Si observamos el gráfico 2.1, la evolución temporal de los precios de alquiler y venta de vivienda en España durante la crisis de 2008 siguieron una tendencia decreciente. En los precios de los alquileres observamos una tendencia decreciente más clara mientras que en los precios de venta podemos ver posiblemente los efectos que tuvo la ligera recuperación que se produjo en 2010 en Europa. Este pico en los precios de venta podría ser fruto de una falsa salida de la crisis que dura hasta aproximadamente 2011 cuando la economía vuelve a desplomarse alcanzando el mínimo en 2013 con una tasa de variación interanual del -3.5% que podemos observar en el gráfico 2.2.

Atendiendo al índice del idealista, vemos en 2014, los precios de alquiler y venta comienzan a divergir. La falta de recursos y el alto desempleo podrían explicar que se disparara el precio del alquiler de la vivienda, junto con el endurecimiento de las hipotecas consecuencia de la crisis bancaria (García Montalvo, 2019).

Desde 2014, los precios de los alquileres de la vivienda se incrementan considerablemente llegando en poco más de dos años al nivel “precrisis” de 2007, y mantiene tendencia creciente firme hasta el último dato en 2019.

Por el contrario, los precios de venta de vivienda siguen disminuyendo hasta 2017 que comienzan a aumentar poco a poco.

Gráfico 2.1. Evolución de los precios del alquiler en España



Fuente: Banco de España

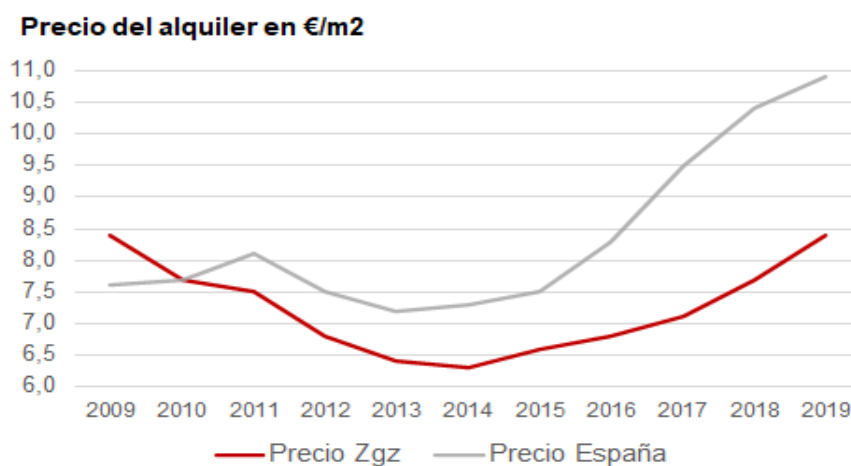
Gráfico 2.2. Tasa de variación interanual PIB



Fuente: Banco de España

En este contexto, vamos a analizar qué ocurre en Zaragoza con los precios del alquiler. A través del portal El Idealista se puede observar la evolución que ha seguido el precio del metro cuadrado en Zaragoza. Como referencia, también se observa la evolución a nivel nacional.

Gráfico 2.3. Comparación del precio del alquiler Zaragoza - España



Fuente: Elaboración propia

Lo primero que podemos ver en el gráfico 2.3 es que los precios de Zaragoza son claramente inferiores a los de España desde 2010. Con respecto al gráfico 2.1, coincide con que ambos experimentan un rápido crecimiento en 2015 con una pendiente más acusada para el caso español. Los precios para Zaragoza parecen estar sujetos a menos variación que los españoles, aunque siguen la misma senda.

También podemos profundizar y observar cómo fluctúan los precios en distintas zonas de Zaragoza. La tabla 2.1 presenta la tasa de variación anual del precio por metro cuadrado en distintos barrios de Zaragoza.

Tabla 2.1. Variación anual 2020 del precio

Barrios	Variaciones
Actur	1,2%
Almozara	0%
Arrabal	1,4%
Caso Histórico	5,4%
Centro	3%
Delicias	7,7%
Las Fuentes	1%
San José	3,1%
Valdespartera	0%
Oliver-Valdefierro	0%
Romareda	6,2%
Parque Goya	0%
Torrero-La Paz	8,7%
Universidad SF	13,2%
Zaragoza	6,2%

Fuente: Idealista

La tasa de crecimiento interanual corresponde al periodo 2019-2020 y muestra la variación en el precio del metro cuadrado que han experimentado los barrios de la ciudad de Zaragoza. Las zonas con mayores incrementos del precio han sido Casco Histórico, Universidad de San Francisco y la Romareda. Junto con el Centro, son los barrios más caros de Zaragoza.

Las zonas como Almozara, Valdespartera, Oliver-Valdefierro y Parque Goya se han mantenido constantes este último año. Y en términos agregados, el precio del alquiler en Zaragoza se incrementó 6.2 puntos porcentuales en el último año.

Según Montalvo (2019), una de las causas del incremento de los precios en España que popularmente deducimos es el crecimiento de los inversores a gran escala y de los apartamentos turísticos, aunque no existen evidencias empíricas suficientes como para afirmar tal cosa ya que los grandes propietarios de viviendas no representan más del 5% y no existen datos suficientes que defiendan la relación entre los pisos turísticos y el incremento de los precios del alquiler. Concretamente Zaragoza, no se caracteriza por el turismo, pero sí que puede considerarse ciudad de paso al estar cerca de ciudades como Barcelona, Madrid, País Vasco y Valencia las cuales sí que reciben un alto número de turistas al año, según la encuesta de ocupación hotelera realizada por el INE. No obstante, Zaragoza es una ciudad de referencia tanto en hospitales como universidades por lo que la población universitaria puede influir en el precio de los alquileres. Además, Zaragoza cuenta con una alta población inmigrante que también puede influir en los precios del alquiler.

Algunos trabajos analizan el precio de venta o alquiler de vivienda y nos pueden dar una ligera idea de los factores que pueden influir en los precios del alquiler de vivienda.

En Wooldridge (2013), aparecen modelos que tratan de explicar los precios de compra de las viviendas en EE. UU. para el año 1995. No es exactamente el objetivo de este trabajo, pero ofrece cierta inspiración. Las variables explicativas que usa son relacionadas con los metros cuadrados, el número de habitaciones, la distancia a puntos, como a un incinerador de basura o/y la cantidad de contaminación. Evita introducir simultáneamente las variables metros cuadrados y habitaciones por alta multicolinealidad. Concretamente, plantea un modelo semilogarítmico en el que los precios están en logaritmos, además de algunas variables explicativas como la contaminación y la distancia.

Díez Muñoz (2017), plantea un modelo econométrico con el objetivo de explicar los precios del alquiler de la vivienda por metro cuadrado en Barcelona para T barrios distintos tomando como variables explicativas el precio de venta de la vivienda medio por metros cuadrados, el salario medio que se observa en cada barrio y el precio medio por plaza en viviendas de uso turístico en cada barrio. En este, caso, el enfoque es distinto ya que este trabajo se centra en el impacto que genera el turismo en una ciudad como Barcelona.

No obstante, puede haber otros factores que determinen la calidad o/y las preferencias del comprador. Seguí Cortés (2017), clasifica las posibles características que pueden influir en el precio de venta de una vivienda para el caso de la ciudad de Altea en Alicante. Las divide en cuatro categorías:

1. Estructura de la vivienda, donde se encuentran factores como la superficie construida, la antigüedad de la vivienda y el número de baños y habitaciones.
2. Características internas de la vivienda, como su orientación solar, si tiene o no terraza, aire acondicionado y/o certificado energético.
3. Características externas de la vivienda, referidas a si tiene ascensor, garaje, trastero, piscina comunitaria etc.
4. Entorno natural que engloba las vistas de las que dispone la vivienda y su proximidad al mar.

Para el caso de Zaragoza, tendremos en cuenta algunos aspectos de estos trabajos ya que la zona en la que está situado el piso puede ser importante en el proceso de decisión además de las características de la vivienda.

Ya sea por como esté comunicada, por como de antigua sea, la distancia a los campus universitarios, el nivel de delincuencia que haya o si tiene o no zonas verdes. Sobre estos factores hemos elegido variables que puedan aproximarse como la distancia al tranvía o la distancia a Plaza España, que podemos considerarla un punto céntrico de la ciudad. Mediante la agrupación de los distritos podemos señalar qué zonas recogen las características que interesan como las mencionadas anteriormente.

En el caso de las características de la vivienda están muy relacionadas con las preferencias de cada arrendatario. El número de habitaciones, de baños, la luminosidad que tenga, si

tiene ascensor o no, terraza o si puede entrar a vivir al momento. Puede que para algunos arrendatarios este último atributo sea positivo porque pueden mudarse inmediatamente o puede que sea negativo, ya que, dependiendo de la antigüedad, algunos pisos que tienen potencial pierden compradores. También puede ocurrir que el arrendatario quiera elegir sus muebles por lo que valoraría que no estuviera amueblado.

Para el caso de la luminosidad que tiene el piso, podemos aproximarnos mediante la variable planta que recoge este aspecto además de otros como el ruido o la seguridad. Una aproximación más correcta hubiera sido tener en cuenta el número de plantas del edificio y realizar una ratio número de planta partido por total de número de plantas, pero dicho dato no se ha podido obtener.

Y por último, un factor que puede ser interesante tener en cuenta es si los precios han sido decididos por particulares o por una agencia. En algunos casos, si son fijados por particulares pueden estar inflados por diversos motivos como que quieran obtener una mayor renta de su inmueble. No obstante, puede ocurrir que estén más bajos que los que fijan las agencias ya sea porque quieren hacer una oferta más atractiva o porque lo alquilen a familiares o allegados y prefieran favorecerles fijando un precio más bajo.

Todos estos factores los tendremos en cuenta posteriormente en la construcción de los modelos econométricos.

Metodología

En este trabajo usaremos la econometría como herramienta metodológica. Son muchas las definiciones que podemos encontrar sobre Econometría. Según Wooldridge (2013), la econometría se apoya en el desarrollo de métodos estadísticos que sirven como herramienta para estimar relaciones económicas, probar teorías y evaluar e implementar políticas públicas y de negocios. Para Otero (1978), “La econometría se ocupa de la cuantificación de los fenómenos económicos y de la verificación de las teorías económicas, haciendo uso para ello de las estadísticas económicas y aplicando a tal fin métodos especiales de inferencias estadística.”

Por otro lado Judge et al. (1982) establece que “La econometría, utilizando la teoría económica, economía matemática e inferencia estadística como fundamentos analíticos y los datos como fuente de información, proporciona a la Ciencia Económica una base para: modificar, refinar o posiblemente refutar las conclusiones contenidas en el cuerpo de conocimientos, conocido como teoría económica y conseguir signos, magnitudes y proporciones fiables acerca de los coeficientes de las variables en las relaciones económicas, de modelo que esta información puede servir de base para la toma de decisiones y la elección.”

Definida de una u otra forma, la econometría parte de la realidad para juzgar la fiabilidad de cualquier relación entre variables ya que emplea el análisis y la recolección de datos no experimentales. Tal como señala Wooldridge, nos podemos encontrar con tres tipos de datos: de corte transversal, series temporales y datos de panel. Los datos de corte transversal recogen una muestra de individuos o cualquier unidad que se toma en un momento del tiempo determinado. Puede ocurrir que los datos no se correspondan al mismo momento del tiempo. Para el caso de los datos de corte transversal puros, estas diferencias temporales son prescindibles. Otras de sus principales características son que el orden de los datos no va a tener ninguna repercusión en su análisis. Además, éstos están estrechamente relacionados con la microeconomía aplicada ya que pueden probar hipótesis micro y/o evaluar políticas económicas (Wooldridge, 2013). Por otro lado, los datos pueden ser de series temporales, que reúnen la evolución de una variable en función del tiempo. Por último, los datos de panel combinan ambas dimensiones.

A diferencia de las ciencias naturales como la Física, Química o Biología en las que se obtienen los datos mediante experimentos controlados, la economía estudia fenómenos sociales por lo que la producción de datos es más difícil de obtener (Pulido, 1983). Por ello, el objetivo de la econometría es confrontar los hechos económicos observados con las hipótesis que se plantean sobre ellos. Con este fin, se construyen los modelos que son una representación simplificada mediante símbolos matemáticos de cierto conjunto de relaciones entre k variables. Una forma genérica de representación de estos modelos sería:

$$y = f(x_1, x_2, x_3, \dots, x_k)$$

Esta definición se acerca más al concepto de modelo económico, mientras que el econométrico presenta una forma funcional definida, las variables que van a influir en la explicación están también definidas, y hay un elemento aleatorio que recoge errores de medida y efectos de factores individualmente irrelevantes.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u$$

Cualquier modelo econométrico se puede dividir en dos partes, la parte sistemática y la parte aleatoria. La sistemática recoge las relaciones y efectos entre los regresores y la variable a explicar. Tanto la variable dependiente como los regresores son magnitudes cuyos correspondientes conjuntos de valores son resultado de la observación de una muestra aleatoria simple de tamaño T (Wooldridge, 2013). La parte aleatoria es el último término denominado “ u ” o perturbación. Como hemos señalado, en la perturbación podemos encontrar los errores de medición que se hayan podido producir y factores que individualmente son irrelevantes en la explicación de la variable independiente, Trívez (2004). Al ser un elemento aleatorio, sigue una distribución de probabilidad, es decir, que se puede observar cómo se reparte la probabilidad sobre los valores que toma la variable. La aleatoriedad de esta perturbación hace aleatoria a la variable dependiente, y ambas siguen la misma distribución de probabilidad.

La metodología econométrica se estructura en cuatro etapas: especificación, estimación, validación y explotación del modelo.

En la etapa de especificación se construye el modelo en términos probabilísticos, es decir, modelizamos el comportamiento de una variable “ y ”. En la medida de lo posible suele intentar modelizarse como una función lineal (o linealizable) de una serie de variables explicativas y una perturbación aleatoria. Al ser un modelo clásico lineal, vamos a trabajar con una serie de supuestos. En concreto, los regresores no serán considerados estocásticos, por lo que la parte sistemática y la aleatoria deben ser independientes. Además, las variables explicativas tienen que ser linealmente independientes, es decir, debe darse ausencia de multicolinealidad perfecta. El número de parámetros que tenga el modelo es k y T es el número de observaciones muestrales de las que disponemos para cada una de las variables del modelo. T tiene que ser estrictamente mayor que el número de parámetros k . Sobre estos parámetros no hay ninguna restricción.

Las hipótesis de la perturbación aleatoria implican que su esperanza matemática es cero, es decir, que el conjunto de las variables individualmente irrelevantes que hay dentro de la perturbación se compensan en promedio; además todas las perturbaciones tendrán la misma variancia, no presentan autocorrelación por pares, siguen una distribución normal. Estos supuestos se resumen en la siguiente expresión.

$$u \sim iid N(0, \sigma^2 I_T)$$

Una vez elegidas las variables explicativas, se formula una relación existente ya sea lineal o linealizable. Por linealizable nos referimos a formas funcionales no lineales, como el uso de logaritmos neperianos o formas cuadráticas, que mediante herramientas como cambios de variables pueden seguir manteniendo la linealidad. En ocasiones, los modelos pueden presentar formas no lineales y no linealizables pero este caso no lo contemplamos.

Los factores que potencialmente explican el precio pueden ser cuantitativos o cualitativos. Para cuantificar los factores cualitativos se utilizan las variables ficticias o dummies, que adopta valores cero y uno. Mediante el ejemplo que ofrece Wooldridge, explicamos su funcionamiento. Sea el modelo:

$$\omega_i = \beta_0 + \beta_1 edu_i + \beta_2 mujer_i + u$$

donde ω es el salario, edu es la educación de cada individuo y $mujer$ es una variable ficticia que da valor 1 cuando el individuo es mujer y 0 cuando es hombre. En este caso la variable ficticia se introduce de forma aditiva pero posteriormente explicaremos que se puede introducir en el modelo de forma multiplicativa.

Veremos que si el individuo es mujer, el salario esperado será $(\beta_0 + \beta_2)$, lo que nos permite saber si por el hecho de ser mujer, el salario va a ser menor, mayor o igual al de un hombre.

Si introducimos la variable de forma multiplicativa, tendremos:

$$\omega_i = \beta_0 + \beta_1 \text{edu}_i + \beta_2 \text{mujer}_i + \beta_3 (\text{edu} * \text{mujer})_i + u$$

Mediante β_3 podemos cuantificar el efecto de la educación sobre el salario cuando el individuo es mujer y observar si la educación de una mujer influye sobre el salario más, menor o lo mismo que la de un hombre.

Como podemos observar en el ejemplo, aunque hay dos ficticias, hombre y mujer, la variable hombre se queda fuera, para evitar problemas de multicolinealidad exacta, que impedirían estimar los parámetros. Esta variable recibe el nombre de variable de control. Lo mismo ocurre cuando tenemos más de dos categorías, habrá que introducir tantas ficticias como categorías menos una para evitar la trampa de las ficticias.

En la etapa de estimación cuantificamos los parámetros de posición (β 's) y de dispersión (σ^2) mediante dos métodos: MCO y MV. Cuando el modelo es lineal, el método MCO coincide con MV.

El método Mínimos Cuadrados Ordinarios o MCO consiste en minimizar la suma de las diferencias al cuadrado entre los valores reales y estimados de la variable endógena, es decir, la suma de los cuadrados de los residuos.

Los estimadores obtenidos son una combinación lineal de los valores de las variables endógenas. Si se mantienen los supuestos comentados anteriormente, también cumplirán la insesgadez, por lo que:

$$E(\hat{\beta}) = \beta$$

Asimismo, serán estimadores óptimos (ELIO). Además, de ser insesgado alcanzará la cota de Cramér-Rao, es decir, su varianza es la mínima que puede alcanzar un estimador de un parámetro dadas unas condiciones de regularidad, por lo que será eficiente (Pulido, 1983). Las propiedades que presenta para muestras grandes (o propiedades asintóticas) son insesgadez asintótica y consistencia. Las condiciones suficientes para que un estimador sea consistente son que el límite del valor esperado del estimador de posición cuando T tiende a infinito sea el valor del parámetro verdadero y que el límite de la matriz de varianzas y covarianzas cuando T tiende a infinito sea cero. Es importante señalar que, si el estimador es insesgado, también será insesgado asintóticamente pero no al revés.

Las propiedades del parámetro de dispersión también dependen del tamaño de la muestra. Para muestras grandes, son similares que para los β 's, asintóticamente insesgados y consistentes mientras que para muestras finitas sólo mantiene la insesgadez y no son lineales en ninguno de los casos.

El método Máximo Verosímil o MV consiste en maximizar la función logarítmica de verosimilitud y es el método de estimación más general ya que no tenemos que garantizar la linealidad en nuestro modelo. Como señala Pulido (1983), "Los estimadores máximo-verosímiles de los parámetros del modelo son aquellos que son más probables dada la distribución de los datos muestrales y su implicación sobre la función de densidad conjunta." Bajo condiciones muy generales en los modelos no lineales, el método MV garantiza el cumplimiento de las condiciones mínimas deseables, fundamentalmente la

consistencia. Pero si el modelo es lineal, los estimadores de posición MCO y los MV coinciden. No ocurre lo mismo para el parámetro de dispersión. La estimación MV del parámetro de dispersión es sesgada, pero en muestras grandes es asintóticamente insesgado y consistente.

La etapa de validación sirve para saber si nuestro modelo es fiable y si de verdad explica nuestra variable objeto de estudio. En esta etapa se considera una amplia batería de contrastes de hipótesis, y otras medidas escalares de bondad de ajuste. Muy importantes los contrastes de significatividad individual y conjunta de las variables explicativas del modelo.

Los contrastes de significatividad individual o t-ratios consisten en plantear:

$$H_0: \beta_j = 0$$

$$H_A: \beta_j \neq 0$$

siendo H_0 la hipótesis nula y H_A la alternativa. Si no rechazamos H_0 , concluimos que el regresor que acompaña a β_j , es decir X_j , no es significativo. Si embargo si rechazamos H_0 , X_j sí está explicando a la variable dependiente.

Las hipótesis se rechazan o no para un nivel de significatividad que se fija previamente. Este nivel es la probabilidad de cometer un error tipo I, lo que quiere decir que rechazamos una hipótesis nula siendo cierta. Por lo que cuanto menor sea, mejor. Suele ser 5% pero también puede ser 1% cuando no admitimos casi errores y 10% si somos más laxos.

Cuando queremos contrastar la significatividad conjunta planteamos:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_A: \text{Alguno} \neq 0$$

En la H_0 igualamos a cero todos los parámetros menos la constante. La hipótesis alternativa es que alguno de ellos sea distinto de cero, pero tampoco sabemos concretamente cual. Si no rechazamos la hipótesis nula, las variables explicativas de nuestro modelo no influyen en la variable dependiente, por lo que ese modelo planteado no sería válido.

El coeficiente de determinación o R^2 es una medida escalar de bondad de ajuste. Consiste en analizar la variación muestral de la variable endógena mediante descomposiciones de forma que se llega a esta expresión:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \rightarrow ST = SE + SR$$

siendo ST la suma total, SE suma esperada y SR la suma residual.

El R^2 es el cociente entre la SE y ST (Trívez, 2004).

Sin embargo, asegurarnos de que se cumplen las hipótesis inicialmente asumidas es importante para que los contrastes de significatividad y otros sean válidos. Por ejemplo, la posible existencia de multicolinealidad aproximada puede afectar a las varianzas de los estimadores siendo esta mayor, por lo que al calcular los t-ratios son más pequeños y termina afectando a la significatividad. Además, que la perturbación aleatoria no siga una distribución normal, covarianza constante y no autocorrelación, también invalida la inferencia estadística.

En este sentido, primero deben verificarse las propiedades de la perturbación, heterocedasticidad (varianza constante), no autocorrelación y normalidad para que los t-ratios y los contrastes de significatividad conjunta (F) sean válidos.

Posteriormente, conviene saber si se cumplen las hipótesis básicas en la parte sistemática, ausencia de multicolinealidad perfecta y si la forma funcional considerada es o no adecuada. Si la forma funcional es incorrecta puede afectar al sesgo de los estimadores. Esta etapa puede ser muy extensa y depende mucho del tipo de datos que tengamos sobre todo en la parte aleatoria. Para datos de corte transversal, no es necesario contrastar la no autocorrelación de la perturbación debido a que los datos corresponden a un periodo de tiempo determinado mientras que para series temporales la no autocorrelación es muy importante. Con respecto a la heterocedasticidad debemos asegurarnos de que no provoque problemas, pero sus efectos negativos tienen una solución fácil, que es la estimación con desviaciones robustas a heterocedasticidad. Aquellos modelos que superen la batería de contrastes podemos considerarlos candidatos a explicar nuestra variable. Entre ellos, seleccionaremos el más adecuado.

La selección es la última etapa de la metodología y depende de cuál sea el interés de explotación. Ésta puede abarcar uno o diferentes fines ya sea el análisis estructural, predicción o/y simulación (Trívez, 2004).

El análisis estructural se centra en el signo y el tamaño de los parámetros mientras que la predicción busca determinar los valores futuros de la variable objeto de estudio, conociendo los valores del resto de variables. Y la simulación se basa en generar trayectorias de las variables a explicar ensayando cambios en los regresores.

Los criterios de selección más utilizados cuando nuestro fin es el análisis estructural son SBIC y AIC. El modelo que tenga el valor más pequeño será el mejor.

El SBIC favorece al modelo más parsimonioso, es decir, que con menos variables explique lo mismo que otro con más. En cambio, el AIC favorece la capacidad predictiva del modelo.

Los criterios más usuales para seleccionar el modelo mejor cuando nuestro objetivo fundamental es predecir son el Error Cuadrático Medio Ponderado (ECMP) y el Error Absoluto Medio Ponderado (EAMP). Ambos se expresan en porcentaje y el modelo que tenga menor porcentaje será mejor.

En definitiva, debemos intentar combinar estos criterios, si nuestros objetivos son ambos, esto es, análisis estructural y predicción.

Variables y datos

La variable que pretendemos explicar es el precio del alquiler mensual de las viviendas en la ciudad de Zaragoza, medida en euros. En concreto, esta variable corresponde al mes de octubre de 2019.

Teniendo en cuenta los estudios previos y las características de Zaragoza, proponemos una serie de factores que potencialmente pueden explicar el precio del alquiler de una vivienda. Los podemos dividir en dos grupos: los que hacen referencia a la ubicación de la vivienda y los que se centran en sus características internas y externas.

Con respecto al primer grupo pueden incluirse los siguientes:

- Renta de la zona. Nos permite tener en cuenta el nivel económico de la zona en la que se encuentra la vivienda que se ofrece por lo que podría darnos una aproximación del nivel de delincuencia, antigüedad... Al ser un dato de renta media, puede medirse por persona o por hogar.
- Zona. Independientemente de la renta de la zona, pueden existir otras características que influyan en el precio. No obstante, tanto la renta como la distancia a puntos concretos, que se comenta a continuación, ya están teniendo en cuenta la zona, por lo que es posible que este efecto ya esté recogido y no requiera ninguna variable concreta.
- Distancia a puntos concretos de la ciudad. Al ser Zaragoza una ciudad universitaria, este factor puede recoger la importancia de que una vivienda se encuentre próxima a los campus San Francisco o al CPS, además de otros puntos estratégicos como Plaza España.

El segundo grupo podría incluir los siguientes factores relativos a las características de la vivienda:

- Tamaño de la vivienda. La superficie va a ser influyente en el precio.
- Número de habitaciones. Puede que esté relacionado con la superficie, aunque no está claro en qué sentido.
- Número de baños. Podemos intuir que tener más de un baño en una vivienda puede afectar sustancialmente al precio.
- Altura. A través del número de planta, podemos tener en cuenta la luminosidad o el ruido. A priori, más luminosidad hará aumentar el precio mientras que mayor ruido lo reducirá. Sin embargo, intuimos que el aumento en la planta dependerá de si el edificio tiene o no ascensor.
- Ascensor. Es muy probable que su efecto sobre el precio sea positivo.
- Estar amueblado. Es un factor que en principio no sabemos si va a influir en el precio, y si lo hace, desconocemos el sentido. Probablemente de las preferencias del arrendatario y posiblemente de la renta media de la zona.

Los datos sobre todos estos factores se han obtenido a partir de la plataforma Idealista, así como información adicional de otras fuentes (INE y Ayuntamiento de Zaragoza) que nos han permitido completar las características concretas de los distritos de la ciudad.

Parte de estos datos aparecen en Segura (2019).

La tabla 4.1. recoge los factores que hemos comentado anteriormente, indicando el nombre del factor explicativo, y la variable con la que podemos medirlo. En algunos casos nos planteamos varias alternativas para medirlo, que pueden ser o no excluyentes. Disponemos de una muestra de 790 viviendas

Tabla 4.1. Variables concretas definidas

Factor	Variable	Cuantificación/medida de la variable
Renta de la zona	renta_persona	Ingresos anuales (euros) que percibe un individuo en 2016
	renta_hogar	Ingresos anuales (euros) que percibe un grupo de personas que comparte un mismo domicilio en 2016
Distancias a puntos concretos	d_SF	Distancia en kilómetros que existe entre el campus San Francisco y el piso
	d_CPS	Distancia en kilómetros que existe entre el campus Río Ebro y el piso
	d_T	Distancia en kilómetros que existe hasta el tranvía
	d_PE	Distancia en kilómetros que existe entre Plaza España y el piso
Agencia	agencia	Asigna el valor 1 a los pisos cuyo precio lo haya fijado una agencia inmobiliaria y el valor 0 para los que el precio lo fijen particulares
Tamaño	metros	Superficie del piso medida en metros cuadrados
Altura	planta	Nivel al que se encuentra el piso
Amueblado	mueble	Asigna el valor 1 a los pisos que se alquilen con muebles y valor 0 a los pisos en los que no hay muebles
Ascensor	ascensor	Asigna valor 1 a los pisos que tienen ascensor y valor 0 a los que no
Número de baños	baños	Asigna valor 1 a los pisos que tienen más de un baño y 0 a los que tengan sólo un baño
Número de habitaciones	habitaciones	Número de habitaciones que tiene el piso

Esta tabla recoge los factores potenciales inicialmente propuestos, aunque alguno de ellos modifiquemos la forma de medirlos. Por ejemplo, para la variable renta, puede interesarnos medirla por tramos o las variables de distancia podemos considerarlas como un factor cualitativo cerca/lejos todo ello mediante el uso de variables ficticias.

La clasificación de distritos realizada por el INE que nos ha permitido obtener información sobre la renta aparece recogida en la tabla 4.2.

Tabla 4.2. Distritos de Zaragoza

1	CASCO HISTÓRICO
2	CENTRO
3	DELICIAS
4	UNIVERSIDAD-CASABLANCA-DISTRITO SUR
5	SAN JOSÉ
6	LAS FUENTES- CARTUJA BAJA
7	LA ALMOZARA
8	MIRALBUENO-OLIVER-VALDEFIERRO
9	TORRERO LA PAZ
10	ACTUR-ARRABAL
11	JUSLIBOL-MONTAÑANA- MOVERA
12	GARRAPINILLOS-ALFOCEA-MONZALBARBA

A partir de la información sobre el distrito en el que se encuentra la vivienda, podemos construir otros factores cualitativos que podrían resultar importantes como, la antigüedad de la zona, la densidad de la población, número de habitantes, edad media de la zona y nivel de estudios medio. La cuantificación de estos factores nos lleva a definir variables ficticias, que aparecen en la tabla 4.3.

Tabla 4.3. Factores cualitativos de zona y sus correspondientes variables ficticias

Factor	Variables ficticias	Descripción
Antigüedad de la zona	D1_Antig	Asigna el valor 1 a los distritos más antiguos que son: 1, 3, 5, 9
	D2_Medi	Asigna el valor 1 a los distritos que no son ni muy antiguos ni muy nuevos: 4, 7, 8, 11
	D3_Nuev	Asigna el valor 1 a los distritos que son nuevos: 2, 6, 10
Número de habitantes	Poblac	Asigna el valor 1 a los distritos en los que vive mayor número de personas independientemente de la superficie de la zona: 3, 10, 5
	Pobla_baja	Asigna el valor 1 a los distritos en los que vive menor número de personas independientemente de la superficie de la zona: 1, 2, 4, 6, 7, 8, 9, 11, 12
Densidad de población	D_desnbaja	Asigna el valor 1 a los distritos que tienen menor número de personas por superficie: 11, 12, 9
	D_densmedia	Asigna el valor 1 a los distritos que tienen densidad media: 4, 5, 6, 7, 8, 10
	D_densalta	Asigna el valor 1 a los distritos que tienen mayor número de personas por superficie: 1, 3, 2
Edad media de la población	D_viej	Asigna el valor 1 a los distritos en los que su población está más envejecida: 2, 3, 5, 6
	D_edadmed	Asigna el valor 1 a los distritos en los que la mayoría de su población está en edad media: 1, 7, 9, 10, 11, 12
	D_jov	Asigna el valor 1 a los distritos en los que su población es más joven: 4, 8
Nivel de estudios de la población	D_sinestud	Asigna el valor 1 a los distritos en los que tienen mayor número de personas que no superaron la escuela primaria: 8, 9, 6
	D_estudios	Asigna el valor 1 a los distritos en los que su población ha superado la escuela primaria pero no han llegado a obtener estudios superiores: 8, 9, 6
	D_supe	Asigna el valor 1 a los distritos en los que tienen mayor número de personas con estudios superiores: 2, 4

Modelos estimados

La forma más general de las especificaciones que podemos considerar viene dada por la siguiente expresión:

Precio = f(renta, distancias, densidad_de población, antigüedad de la zona, edad media de la zona, planta, ascensor, tamaño, baños,)

A partir de la expresión anterior pueden especificarse modelos lineales o no lineales. Estos últimos consideraremos que son linealizables. Así, si tanto el precio como los regresores se introducen con sus valores en niveles, entonces el modelo es claramente lineal. Por ejemplo:

$$Precio = \beta_1 + \beta_2 renta + \beta_3 planta + \beta_4 metros + \dots + u_i$$

Por el contrario, si los regresores están en logaritmos, mientras que la variable dependiente precio está en niveles, el modelo teórico original es semilogarítmico.

$$Precio = \beta_1 + \beta_2 \log(renta) + \beta_3 planta + \beta_4 metros + \dots + u_i$$

También es semilogarítmico si la variable dependiente está en logaritmos, pero los regresores están en niveles. En este caso el parámetro que acompaña al regresor se interpreta como una semielasticidad. Es lo que ocurre en el siguiente modelo:

$$\log(Precio) = \beta_1 + \beta_2 renta + \beta_3 planta + \beta_4 metros + \dots + u_i$$

Finalmente, si el precio y el regresor están en logaritmos, el correspondiente parámetro representa una elasticidad.

Las estimaciones inicialmente consideradas, a partir de las cuales realizaremos las modificaciones necesarias, se presentan en la tabla 5.1 y 5.3. La tabla 5.1 corresponde a estimaciones de modelos estrictamente lineales, es decir, ni la variable dependiente ni ninguna explicativa está en logaritmos. La tabla 5.2 muestra los contrastes más relevantes sobre la validación de los modelos presentados en la tabla 5.1. En la tabla 5.3 se presentan las estimaciones de los modelos con la variable dependiente en logaritmos, y la validación de estos modelos aparecen en la tabla 5.4.

Hay que tener en cuenta que las variables de las distancias pueden estar relacionadas entre sí. Lógicamente, la proximidad a los campus, San Francisco y CPS pueden estar relacionadas linealmente entre sí, así que evitaremos usar simultáneamente d_{SF} y d_{CPS} .

Tabla 5.1. Estimaciones lineales iniciales. Variable dependiente precio en niveles.

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
const	164,115***	117,04***	100,45***	159,946***	118,11***	219,81***
renta_persona		0,0151***	0,0145***		0,014***	
renta_hogar	0,00438***			0,0043***		0,0077***
d_SF	-7,161					
d_CPS		-6,89				
d_PE				-0,754	0,9163	
d_T	7,585	7,883*	1,9875			2,69
agencia	6,227	5,62				
metros	3,715***	3,697***	3,70***	3,723***	3,7515***	
planta	2,71	2,37	2,449	2,77	2,7418	-0,3346
mueble	12,357	12,099	12,22	11,063		22,06
ascensor	92,752***	80,313***	81,738***	93,51***	82,638***	-44,288*
baños	107,64***	95,934***	96,432***	105,862***	96,27***	174,15***
habitaciones	-2,579	3,503	3,5878	-1,946		86,19***

Tabla 5.2. Validación de las estimaciones lineales iniciales.

		Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Colineali (FIV)	r_persona		1,342	1,248		1,186	
	r_hogar	1,233			1,221		1,132
	d_SF	16,576					
	d_CPS		8,128				
	d_PE				1,044	1,085	
	d_T	16,614	8,53	1,089			1,043
	agencia	1	1				
	metros	2,34	2,278	2,274	2,338	1,544	
	planta	1,085	1,084	1,084	1,084	1,068	1,081
	mueble	1,068	1,064	1,059	1,060		1,060
	ascensor	1,211	1,224	1,221	1,209	1,189	1,067
	baños	1,629	1,345	1,643	1,622	1,447	1,550
	habitac	2,174	2,205	2,202	2,169		1,48
Reset		1,706 (0,182)	1,368 (0,255)	1,706 (0,182)	2,1732 (0,115)	1,568 (0,209)	24,33 (0)
Heteroc	BP	1201,13 (0)	1190,095 (0)	1205,31 (0)	1257,95 (0)	1138,41 (0)	785,95 (0)
	White	277,57 (0)	257,102 (0)	240,962 (0)	264,977 (0)	215,6 (0)	247,473 (0)
R ²		0,6446	0,6592	0,6583	0,644	0,6574	0,4697
Normalid	JB	5539,94 (0)	5706,72 (0)	5929,04 (0)	5439,54 (0)	5870,41 (0)	16497,4 (0)

Los altos Factores de Inflación de la Varianza (FIV) que podemos observar en la tabla son los referidos a las distancias cuando las introducimos simultáneamente en un mismo modelo y entre las variables *baños*, *metros cuadrados* y *habitaciones*. Como podemos

ver en el modelo 6, aun existiendo multicolinealidad entre baños y habitaciones, ésta última es significativa ya que no hemos considerado los metros cuadrados. La multicolinealidad entre estas tres variables afecta a la significatividad de *habitaciones* cuando la introducimos simultáneamente con *metros cuadrados*.

Además, los FIVs de variables como *agencia*, *mueble* y *planta* están muy próximos a 1.

En la misma tabla, presentamos los contrastes básicos de validación, en concreto el contraste sobre la forma funcional, homocedasticidad y normalidad. En cada celda aparece el valor del estadístico y debajo entre paréntesis su correspondiente p-valor.

La hipótesis nula del contraste de Reset es que la forma funcional del modelo es correcta. Rechazaremos la hipótesis nula cuando el estadístico F sea mayor que el valor de tablas que tengamos, dado el número de grados de libertad que siga la distribución F de Snedecor (Trivez, 2004). Si el estadístico es menor que el valor de tablas, no rechazamos la hipótesis nula lo que implica que la forma funcional es correcta. Del mismo modo, un p-valor superior al nivel de significación indica que no se rechaza la hipótesis nula.

Los contrastes que utilizamos para detectar la heterocedasticidad en un modelo son el contraste de Breusch-Pagan y el de White. Ambos contrastes comparten la misma hipótesis nula pero no hipótesis alternativa. La hipótesis nula es que el modelo no presenta heterocedasticidad, es decir, el modelo es homocedástico. El contraste de Breusch-Pagan lo podríamos calcular manualmente estimando el modelo por MCO y obteniendo los residuos al cuadrado (\hat{u}_i^2). Calcularíamos la estimación máximo verosímil del parámetro de dispersión (σ^2) y lo dividiríamos por los residuos al cuadrado. Este cociente va a ser la variable dependiente en nuestra regresión auxiliar, la cual va a representar una relación lineal entre esta variable y las variables del modelo original (Trivez, 2004).

$$\hat{u}_i^2 / \sigma^2 = \alpha_1 + \alpha_2 X_{2i} + \dots + \alpha_p Z_{pi} + \varepsilon_i$$

De la regresión, calculamos su suma explicada (SE) y obtenemos el estadístico BP:

BP = $\frac{1}{2} * SE$; sigue asintóticamente una chi cuadrado de orden p-1, siendo p el número de parámetros de posición de la regresión auxiliar.

A diferencia del contraste de BP, el contraste de White es no constructivo, lo que quiere decir que no podemos concretar qué variable está causando la heterocedasticidad. Manualmente podríamos calcular el estadístico W estimando por MCO y obteniendo los residuos al cuadrado. Construimos una regresión auxiliar que relaciona linealmente los residuos al cuadrado con las variables originales del modelo, estas variables al cuadrado y los términos cruzados (Trivez, 2004).

$$\hat{u}_i^2 = \gamma_1 + \gamma_2 X_2 + \dots + \gamma_k X_k + \gamma_{k+1} X_{2i}^2 + \dots + \gamma_{2k-1} X_{ki}^2 + \gamma_{2k} X_{2i} X_{3i} + \dots + \varepsilon_i$$

A partir de la regresión, obtenemos su R^2 y calculamos el estadístico W:

W = $T R^2$; también sigue asintóticamente una chi cuadrado de orden p-1.

$W = TR^2_{as} \sim \chi^2_r$ siendo r el número de regresores de la regresión auxiliar que incluye las variables explicativas, sus cuadrados y los productos cruzados dos a dos.

Los dos contrastes se resuelven de la misma forma. Rechazaremos la hipótesis nula cuando el estadístico es mayor que el valor de tablas, fijando previamente un nivel de significación del 5%. Al rechazar, concluimos que el modelo presenta heterocedasticidad. El contraste de Jarque-Bera (JB) consiste en contrastar si los residuos del modelo siguen una distribución normal. Para ello se plantea la hipótesis nula de que sí se comportan como una normal, por lo que los coeficientes de curtosis y asimetría son cero. Frente a la hipótesis alternativa de que alguno de estos dos coeficientes son distintos de cero.

Como el estadístico sigue una chi-cuadrado, el contraste se resuelve de forma análoga a White y BP.

Mediante los residuos se calculan los coeficientes de simetría y de curtosis que son necesarios para obtener el estadístico y solucionar el contraste (Trivez, 2004). Si el estadístico es menor que el valor de tablas de una chi cuadrado, no rechazamos la hipótesis nula, es decir, los residuos del modelo seguirán una distribución normal. En cambio, si el estadístico es mayor que el valor de tablas se rechazará la hipótesis nula.

Como podemos observar en la tabla 5.1, los modelos M1 a M5 presentan estimaciones más o menos similares, mientras que el modelo M6 se aleja. Descartamos este modelo al no presentar una forma funcional adecuada. Ninguno de los modelos estimados supera los contrastes de heterocedasticidad ni de normalidad.

Tenemos en cuenta que para comparar los β 's, las respectivas variables tienen que presentar la misma forma funcional.

Independientemente de cómo midamos la renta, en los modelos se mantienen más o menos igual. Para el caso de *renta por hogar*, el estimador suele estar en torno a 0,043. Por lo que, en estas primeras estimaciones, que la vivienda se encuentre en una zona con renta más alta va a suponer un plus de estas cantidades en el precio de alquiler.

Ninguna de las variables distancia es significativa por lo que la interpretación de la estimación del parámetro de posición no es válida.

La variación en el precio ante un incremento en 10 metros cuadrados está entorno a los 37€ al mes.

El incremento en el precio del alquiler de una vivienda cuando tiene ascensor está entorno a [80, 93]€.

El incremento en el precio del alquiler de una vivienda cuando tiene más de un baño está entorno a [95, 107]€. Junto con la variable ascensor, son las variables que tienen un mayor peso sobre el precio del alquiler.

La variación en el precio ante el incremento de una habitación adicional es de 86,19€.

Veamos ahora los resultados de las estimaciones cuando la variable dependiente está en logaritmos.

Tabla 5.3. Estimaciones iniciales con variable dependiente en logaritmos.

	Modelo 7	Modelo 8	Modelo 9	Modelo 10	Modelo 11
const	3,62***	2,7974***	5,92***	2,78***	3,804***
l_renta_persona	0,2579***			0,359***	
l_renta_hogar		0,3286***			0,22***
renta_hogar			8,31e-06***		
d_SF	0,0013				
d_CPS				0,0041**	
d_PE			0,0013		
d_T					-3,26e-05
agencia	0,0054	-0,0079	-0,0032	-0,007	
metros	0,0034***				0,0033***
planta	0,0066***	0,0033	0,0035	0,0034	0,0067**
mueble	0,0187	0,0303**	0,0312**	0,028*	
ascensor	0,131**	0,0155***	0,0257	0,004	0,1383***
baños	0,1628***	0,2089***	0,2143***	0,202***	0,1642***
habitaciones		0,092**	0,089***	0,099***	

Tabla 5.4. Validación primeras estimaciones con precio en logaritmos.

		Modelo 7	Modelo 8	Modelo 9	Modelo 10	Modelo 11
Colineali (FIV)	l_r_persona	1,192			1,13	
	l_r_hogar		1,113			1,179
	r_hogar			1,122		
	d_SF	1,080				
	d_CPS				1,029	
	d_PE			1,037		
	d_T					1,039
	agencia	1,008	1,008	1,007	1,008	
	metros	1,536				1,571
	planta	1,071	1,081	1,081	1,080	1,07
	mueble	1,054	1,062	1,064	1,061	
	ascensor	1,195	1,068	1,069	1,080	1,187
	baños	1,480	1,562	1,551	1,580	1,447
	habitac		1,474	1,485	1,492	
Reset		15,372 (0)	6,591 (0,0001)	5,92 (0,0028)	8,706 (0)	16,94 (0)
Heteroc	BP	166,7 (0)	178,28 (0)	197,046 (0)	181,74 (0)	174,47 (0)
	White	111,108 (0)	117,63 (0)	146,022 (0)	126,726 (0)	106,91 (0)
R ²		0,6719	0,5541	0,5534	0,5664	0,6589
Normalid .	JB	405,295 (0)	531,641 (0)	558,428 (0)	513,064 (0)	415,282 (0)

Observando los FIVs, concluimos que la multicolinealidad entre *baños* y *metros cuadrados* no tiene consecuencias negativas. El resto de las variables tienen FIVs muy próximos a uno.

Ningún modelo pasa el contraste de forma funcional de Reset ni los dos contrastes de heterocedasticidad ni el de normalidad.

Con respecto a la renta, sólo podemos comparar los modelos 7, 8, 10 y 11 ya que en el modelo 9 la variable renta está en niveles. En estos modelos, el incremento de un 1% en

la renta por persona provoca un incremento en el precio del alquiler del 0,26-0,36%. Mientras que para un incremento de 1% en la renta por hogar, los precios aumentan un 0,22-0,33%. Para el caso del modelo 9, la semielasticidad de la renta está muy próxima a cero.

La variación porcentual de los precios ante un metro cuadrado adicional es aproximadamente del 0,3%.

Ante un nivel adicional de altura, los precios se incrementan entre un 0,66%.

Los precios se incrementan alrededor de 3% si el piso se alquila amueblado.

Cuando una vivienda tiene ascensor, el precio de su alquiler se incrementa aproximadamente un 14%.

Los precios del alquiler se incrementan cuando la vivienda tiene más de un baño. Este incremento es mayor en el modelo 9 mientras que en el resto se mantiene entre 16 y 21%.

Tanto en logaritmos como en niveles, baños y ascensor tienen pesos mayores.

Una habitación adicional; incrementa el precio en un 9% aproximadamente.

Las variables *renta*, *metros*, *ascensor* y *baños* casi siempre son significativas y los signos coinciden con la teoría económica. La renta (ya sea por persona u hogar), los metros cuadrados, si tiene ascensor y el número de baños influyen positivamente en los precios del alquiler.

Ningún modelo pasa los contrastes de heterocedasticidad. Por lo que en posteriores modelos realizamos estimaciones con desviaciones robustas a heterocedasticidad.

Con la información que hemos obtenido de las estimaciones iniciales, planteamos estimaciones que se recogen en la tabla 5.5. Como hemos visto en la tabla anterior, no es correcta la forma funcional en los modelos semilogarítmicos en los que la variable dependiente está en logaritmos, por lo que estimamos con la variable en niveles.

Tabla 5.5. Estimaciones lineales con desviaciones robustas a heterocedasticidad. Variable dependiente precio en niveles.

	Modelo 12	Modelo 13	Modelo 14	Modelo 15
const	-1518,85***	161,783***	118,32***	-1678,97***
renta_persona			0,014***	
l_renta_persona				210,082***
renta_hogar		0,0044***		
l_renta_hogar	176,25***			
d_T	-0,08	0,106	1,934	
metros	3,715***	3,68***	3,74***	3,77***
planta	2,742	2,81		2,79
ascensor	88,597***	92,625***	87,075***	80,62***
baños	100,29***	103,123***	97,71***	95,5***
Contraste Reset	1,4813 (0,228)	2,156 (0,116)	1,1741 (0,31)	2,23 (0,108)

Todos los modelos pasan el contraste Reset por lo que sí es correcto un modelo con el regresor de la renta en logaritmos. Los valores de los parámetros estimados son más o menos los que nos han salido en las estimaciones iniciales salvo las semielasticidades de las rentas. La interpretación de este estimador es, que ante un incremento en la renta de la zona de un 1%, los precios se incrementan en 1,76€ para el caso del modelo 12 y 2,10€ para el modelo 15.

Modificaciones y mejoras

A partir de los resultados obtenidos, se pueden plantear otras especificaciones: eliminando las variables no significativas, midiendo los factores a través de otras variables u omitiendo variables que están muy relacionadas con otras.

Para llevar a cabo estas modificaciones ha sido necesario tener en cuenta otras variables, fundamentalmente ficticias, así como posibles efectos de interacción entre variables ya incluidas anteriormente.

La tabla 5.6 recoge información sobre cómo se han construido las variables ficticias y qué valores les hemos asignado.

Tabla 5.6. Variables ficticias adicionales

F1_rpbaja	Valor 1 si la vivienda está en zona con renta media por persona entre 0 y 13000€ sin incluir la zona este último
rp media	Valor 1 si la vivienda está en zona con renta media por persona entre 13000€ y 21000€
F2_rpalta	Valor 1 si la vivienda está en zona con renta media por persona mayor que 21000€
F1_rhbaja	Valor 1 si la vivienda está en zona con renta media por hogar entre 0 y 32000€ sin incluir la zona este último
rh media	Valor 1 si la vivienda está en zona con renta media por hogar entre 32000€ y 45000€
F1_rhalta	Valor 1 si la vivienda está en zona con renta media por hogar mayor que 45000€
univer	Asigna el valor 1 a los pisos que se encuentran a menos de 0.9 km del campus San Francisco y el valor 0 para los que se encuentran a más o igual de 0.9km
centro	Asigna el valor 1 a los pisos que se encuentren a menos de 4.4km de Plaza España y el valor 0 para los pisos que se encuentren a más o igual de 4.4km
ascensor_planta	Producto de dos variables, ascensor y planta.
ascensor_metros	Producto de dos variables, ascensor y metros.
habitac_rhbaja	Producto de dos variables, habitaciones y F1_rhbaja.
habitac_rhalta	Producto de dos variables, habitaciones y F1_rhalta.

Las variables ficticias se pueden incorporar al modelo de dos formas diferentes, aditiva y multiplicativa, dependiendo de los efectos que pretendamos recoger en el modelo. Vamos a introducir la variable *ascensor* de forma multiplicativa con la variable *planta* con el objetivo de confirmar nuestras sospechas. Queremos ver si al aumentar un nivel, los precios del alquiler suben siempre y cuando el edificio tenga ascensor. También introducimos la renta por tramos de forma multiplicativa con la variable *habitaciones* para saber si un mayor número de habitaciones supone precios más bajos cuando la zona tiene menor renta. O si la influencia del número de habitaciones sobre el precio es indiferente del nivel de renta de la zona.

Las tablas 5.7.a, 5.7.b y 5.8 presentan los modelos estimados con las modificaciones realizadas, a partir de la información obtenida en las primeras estimaciones. Concretamente en las tablas 5.7.a y 5.7.b aparecen la renta por tramos, el producto entre *ascensor* y *planta*, el producto entre *metros* y *ascensor*, la mayor o menor cercanía al centro y al campus San Francisco, utilizando dummies, y las ficticias que miden distintas características de la zona. En la tabla 5.8 incorporamos la información obtenida de las tablas 5.7.a y 5.7.b, y tenemos en cuenta el efecto de *habitaciones* cuando la renta de la zona es alta o baja. Además, probamos estimaciones con *habitaciones* en vez de metros cuadrados.

Tabla 5.7.a. Estimaciones con desviaciones robustas a heteroscedasticidad I. Variable dependiente precio en niveles.

	M1	M2	M3	M4	M5	M6	M7
cte	456,4***	255,26***	169,28***	266,68***	-1.663,96***	-1677***	355,46***
l_renta_hogar					195,846***	185,429***	
renta_hogar			0,0043***	0,0043***			0,0038***
F1_rhbaja	-25,38*	-31,6**					
F2_rhalta	138***	124,66***					
centro						24,05	30,59*
metros	3,44***	3,883***	3,689***	3,79***	3,811***	3,836***	3,332***
ascensor	-215,5***	98,45***	76,367***	-55,476	91,269***	89,13***	-238,8***
planta	-52,25***			-60,067***			-55,18***
ascensor_planta	56,686***			65,097***			59,77***
ascensor_metros	2,11***						2,214***
baños	53,59***	104,948** *	101,94***	99,262***	104,269***	103,248***	53,26***
D1_Antig	38,66***	68,78***			68,683***	66***	
D3_Nuev		51,531***			54,5368***	48,68***	
Poblac							-43,41***
Contraste Reset	28,77 (0)	1,98 (0,139)	2,078 (0,126)	4,289 (0,014)	2,431 (0,0886)	3,039 (0,0485)	27,06 (0)
Contraste JB	5687,57 (0)	4766,71 (0)	5490,23 (0)	4285,85 (0)	4917,66 (0)	4970,39 (0)	6644,84 (0)
R ²	0,6857	0,6563	0,6449	0,6566	0,6542	0,655	0,6845

Tabla 5.7.b.

	M8	M9	M10	M11	M12	M13	M14
cte	115,74***	396,85***	481,61***	-1754,5***	418,36***	283,58***	355,2***
l_renta_persona				219,32***			
renta_persona	0,0116***						
F1_rpbaja			-75,31***		-85,077***		-95,135***
F2_rpalta			63,905***		72,152***		81,8***
F1_rhbaja		-15,19				-17,475	
F2_rhalta		127,10***				128,52***	
univer	25,637		2,2089	27,0318	-12,309	10,367	
centro			-8,764			26,66	11,04
metros	3,87***	3,879***	3,53***	3,822***	3,3632***	3,82***	3,862***
ascensor	82,679***	-45,38	-219,5***	84,52***	-105,95***	94,74***	89,36***
planta		-56,12***	-49,86***				
ascensor_planta		61,41***	54,525***				
ascensor_metros			2,037***		2,2383***		
baños	98,55***	97,72***		95,301***	55,949***	100,55***	104,28***
D_densbaja	6,6723						
D_densalta	67,6***		65,475***				
D_viej				-7,767			-12,4
D_jov				-49,638**			-54,33***
D_sinestud		-26,87				-41,855*	
D_supe		-2,25				-8,058	
Contraste Reset	2,095 (0,124)	4,93 (0,007)	30,04 (0)	1,59 (0,205)	22,067 (0)	0,7427 (0,476)	1,83 (0,161)
Contraste JB	5375,77 (0)	4159,32 (0)	6451,78 (0)	5039,46 (0)	9038,55 (0)	4974,96 (0)	4927 (0)
R ²	0,6678	0,6604	0,6957	0,6586	0,675	0,65	0,6534

Los modelos en los que aparece el producto *ascensor_metros* no superan el contraste de Reset. Parece que la variable influye en la forma funcional. En los modelos que no aparece, como el 4 y 9, también hay multicolinealidad, pero esta sí que afecta a la significatividad de la variable ascensor. Los modelos 1, 7, 9, 10, y 12 no presentan una forma funcional correcta al 1% de nivel de significatividad.

Además, ninguno de los modelos estimados supera el contraste de normalidad Jarque Bera. En todos los contrastes rechazaremos la hipótesis nula a un nivel de significatividad del 5% por lo que ninguno de los residuos se comporta como una normal.

Tabla 5.8. Estimaciones con desviaciones robustas a heteroscedasticidad II. Variable dependiente precio en niveles.

	M15	M16	M17	M18
cte	436,645***	441,698***	499,513***	503,02***
F1_rhbaja	-46,248***	8,4	28,28	
F2_rhalta	167,054***	-25,73	8,385	-13,65
univer	13,356	17,024		
centro	-6,725	-2,49		
ascensor	-53,376***	-38,659*	-37,1217	-34,98
baños	173,814***	177,74***	179,9***	181,22***
habita_rhbaja		-21,9	-23,9	
habita_rhalta		66,37***	60,537***	71,79***
habitaciones	94,799***	85,27***	82,623***	71,622***
D_densbaja	92,7633***	102,68***		
D_densalta	86,242***	90,22***		
D_sinestud			-56,26**	-55,78**
D_supe			2,456	14,03
Contraste Reset	35,567 (0)	20,815 (0)	14,44 (0)	16,061 (0)
Contraste JB	12655,2 (0)	10963,6 (0)	120599,9 (0)	12095,1 (0)
R ²	0,4734	0,4921	0,4738	0,4699

Las variables sobre distancias a determinados puntos no son significativas en ninguno de los modelos que presentamos en las tablas 5.7.a, 5.7.b y 5.8.

Los modelos de la tabla 5.8 no superan ninguno el contraste de Reset ni de normalidad y sus respectivos R²s se encuentran 20 puntos porcentuales por debajo de los de las tablas 5.7a y 5.7b, donde introducimos los metros cuadrados de la vivienda.

En la tabla 5.9 representamos modelos semilogarítmicos para poder observar si las modificaciones que hemos realizado nos ofrecen un resultado alternativo al obtenido en la tabla 5.3.

Tabla 5.9. Estimación con desviaciones robustas a heterocedasticidad. Variable dependiente en logaritmos.

	M22	M23	M24	M25	M26
cte	6,18***	6,27***	6,19***	5,964***	6,2289***
renta_hogar				6,15e-06***	
F1_rpbaja	-0,1***				
F2_rpalta	0,102***				
F1_rhbaja			-0,0337***		-0,031**
F2_rhalta		0,1064**	0,1145***		0,1402***
univer	0,0268				0,022
centro			0,0002		
metros	0,0035***		0,003***	0,0027***	0,0029***
ascensor	0,0168	-0,0482	-0,1725***	-0,0947**	-0,1798***
planta	-0,044**	-0,0261	-0,0345**		-0,0367**
ascensor_planta	0,0532***	0,0328	0,04178**		0,0448***
ascensor_metros			0,0025***	0,00281***	0,0026***
baños	0,0619***	0,22***	0,1101***	0,109***	0,1122***
habitac_rhalta		0,0368**			
habitaciones		0,084***			
D1_Antig	0,0619***			0,0387***	0,0524***
D3_Nuev	0,04678***				
Poblac					-0,063***
D_densalta			0,0784***		
D_sinestud		-0,077***			
Contraste Reset	11,39 (0)	2,677 (0,0694)	3,619 (0,0272)	6,315 (0,0019)	3,327 (0,0364)

Podemos observar que forma funcional es correcta en los modelos 23, 24 y 26.

Modelos que superan la validación

En las siguientes tablas se presentan las estimaciones de los modelos que superan el contraste Reset. Como hemos visto, en las estimaciones anteriores los modelos no superan el contraste de normalidad Jarque-Bera. Al trabajar con una muestra grande, podemos considerar que la inferencia realizada es válida.

Algunos de los modelos se corresponden con los que aparecen en tablas anteriores mientras que otros están modificados de forma que recojan la máxima información. En total son doce modelos que superan nuestra validación, de los cuales seleccionaremos el que se considere mejor, atendiendo a los objetivos del investigador. También es importante destacar que en el modelo E y L, la variable dependiente está en logaritmos y en el resto de los modelos se encuentra en niveles.

Tabla 5.10.a. Estimaciones con desviaciones robustas a heterocedasticidad.

	M_A	M_B	M_C	M_D	M_E	M_F
cte	176,467***	161,841***	104,96**	320,012***	6,247***	118,474***
renta_persona						0,0117***
renta_hogar	0,00344***	0,0039***	0,0045***	0,0039***		
F2_rhalsa					0,1244**	
univer		39,19**	41,89**			
centro	49,79**	42,154**				
metros	3,762***	3,801***	3,80***	3,739***		3,867***
ascensor	87,39***	92,053***	94,49***	-52,02		83,34***
planta				-57,24**		
ascensor_planta				62,89***		
baños	105,724***	104,248***	105,342***	102,24***	0,2222***	99,484***
habitac_rhalsa					0,031*	
habitaciones					0,086***	
D1_Antig			72,42***			
D3_Nuev			60,654***			
Poblac	-52,4***	-52,815***		-38,93***		
D_densalta						64,12***
D_jov		-59,08***				
D_sinestud	-47,156**			-30,068	-0,09***	

Tabla 5.10.b.

	M_G	M_H	M_I	M_J	M_K	M_L
cte	244,88***	292,08***	274,2***	235,204***	-1657,15***	6,196***
l_renta_hogar					184,416***	
F1_rhbaja	-27,67**					-0,0337***
F2_rhaltura	126,36***	136,38***	114,1***	120,07***		0,1145***
univer	37,553**			37,67**	38,35**	
centro			45,48**			
metros	3,90***	3,8***	3,84***	3,923***	3,83***	0,003***
planta						-0,0345**
ascensor	97,885***	99,11***	91,56***	94,37***	90,155***	-0,1725***
ascensor_planta						0,042**
ascensor_metros						0,0025***
baños	103,423***	102,816***	104,6***	102,013***	102,637***	0,1109***
D1_Antig	74,074***				74,78***	
D3_Nuev	57,79***				60,68***	
Poblac			-57,92***			
D_densalta				82,96***		0,0784***
D_sinestud		-44,8**	-57,824***			

Los parámetros estimados en los modelos anteriores nos permiten ver el efecto que cada uno de los factores explicativos tiene sobre el precio de la vivienda:

- El nivel de renta por hogar que hay en una zona determinada tiene un peso muy pequeño. Ante un incremento de 1000€ en la renta, los precios se incrementan 4€ aproximadamente. Tomando como referencia el modelo G, en las viviendas que se encuentran en una zona rica tienen un incremento en el precio del alquiler de 126,36€. Mientras que, si la zona tiene menos recursos, los alquileres son 27,67€ más baratos. No obstante, también hemos tenido en cuenta otras formas funcionales como en el modelo E y K. En este último, un incremento de 1% en la renta de zona supone un incremento de los precios de 1,84€.
- Las viviendas que se encuentran cerca de la universidad son unos 40€ más caras en su alquiler mientras que el alquiler de viviendas céntricas son aproximadamente 46€ más caras. Siendo Zaragoza una ciudad universitaria, los datos y la teoría económica coinciden.
- Podemos observar que cuando una vivienda tiene ascensor, el precio de alquilarla puede suponer un incremento de 83€ a 99€.

- El modelo E es el único caso en el que aparece el número de habitaciones. Podemos ver que si la vivienda tiene una habitación más en una zona de renta baja y media se incrementa en un 8,6% el precio del alquiler. Por el contrario, si el nivel de renta de la zona es alto, una habitación más en la vivienda supone un aumento de 11,7% en el precio (suma entre 8,6% y 3,1%).

- Según las características que tenga la zona en la que se encuentra la vivienda, vemos que los precios de los alquileres pueden variar considerablemente. Si la vivienda se encuentra en una zona antigua, los precios de su alquiler pueden incrementarse 73€ aproximadamente, mientras que si la vivienda se encuentra en una zona nueva también pueden incrementarse uno 60€. Por lo que comparativamente es ligeramente más cara la zona antigua.

También podemos observar que, si la zona tiene un mayor número de jóvenes, los precios del alquiler son aproximadamente 57€ más baratos. Y, por último, si la zona tiene un alto porcentaje de personas sin estudios, los precios del alquiler pueden ser 30 o 57€ más baratos.

Selección del modelo

Una vez determinados los modelos que superan la etapa de validación, tenemos que elegir cuál es el mejor. Realizamos la selección mediante los criterios de información, como AIC y SBIC, y medidas predictivas, como ECM (Error Cuadrático Medio) y EAPM (Error Absoluto Porcentual Medio). Para obtener estos últimos tendremos que hacer predicción de las cinco observaciones que previamente habíamos sacado de la muestra. Al conocer el valor verdadero, podemos observar si se encuentra dentro o fuera del Intervalo de Confianza (IC) de la predicción. Si está dentro del IC significa que el modelo cumple la Permanencia Estructural (PE), es decir, que la estructura, variables y demás del modelo se mantienen fuera de la muestra. En caso contrario, detectaremos una ruptura estructural y la predicción no será válida.

En la siguiente tabla presentamos los criterios mencionados.

Tabla 6.1. Criterios de selección de los modelos que superan validación.

	AIC	SBIC	ECM	EAPM	IC
M_A	10411,04	10448,41	159,81	15,733%	PE
M_B	10407,82	10449,87	154,57	15,259%	PE
M_C	10402,31	10439,68	158,41	16,094%	PE
M_D	10395,32	10437,37	153,19	14,376%	PE
M_E	-224,8	-196,769	29499,11	18,928%	PE
M_F	10367,27	10395,30	149,8	16,922%	PE
M_G	10395,19	10437,24	170,15	16,381%	Ruptura
M_H	10411,17	10439,20	175,35	16,734%	Ruptura
M_I	10400,17	10437,54	172,92	16,179%	Ruptura
M_J	10374,76	10407,46	166,47	17,455%	Ruptura
M_K	10398,09	10435,46	157,59	16,283%	PE
M_L	-591,88	-545,168	59230,31	19,85%	Ruptura

Seleccionamos el modelo que menores valores presente de los criterios expuestos. Es importante recordar que los modelo E y L están estimados con la variable dependiente en logaritmos por lo que hemos realizado una transformación para poder comparar su capacidad predictiva con los demás modelos. Dicha transformación es la siguiente:

$$P_0 = \exp(\ln P_0 + \frac{1}{2} * \sigma^2)$$

siendo P_0 la variable transformada en niveles y σ^2 la varianza estimada del modelo, igual para todas las observaciones.

Los resultados obtenidos están expuestos en la tabla 6.1, asimismo descartamos los modelos E y L por su baja capacidad predictiva. El modelo que menores valores de AIC, SBIC y ECM es el modelo F y el modelo que menor porcentaje de EAPM (Error Absoluto Porcentual Medio) tiene, es el modelo D.

Todos los modelos que presentan ruptura estructural presentan dicha ruptura sólo en la primera predicción (observación 791). El valor verdadero se encuentra por debajo del límite inferior del IC que hemos obtenido.

Explotación

En el siguiente apartado damos uso a los modelos seleccionados D y F. En primer lugar, realizamos el análisis estructural correspondiente, centrándonos en lo más representativo. En ambos modelos podemos observar que el peso de los metros cuadrados y de los baños es similar. Una vivienda con 10 metros cuadrados más supone un aumento en el precio de 38€ aproximadamente y si la vivienda tiene más de un baño, los precios se pueden incrementar unos 100€. Con respecto a la renta de la zona, vemos que es un efecto débil comparado con el resto de los factores.

Sin embargo, una de las diferencias más destacables es que el modelo D tiene variables que hacen que el precio del alquiler sea más barato (pueda disminuir) mientras que en el modelo F todas las variables añaden valor. Según el modelo D, una zona en la que la haya mucha población y que la vivienda esté más alta en un edificio sin ascensor son factores que van a hacer el precio del alquiler más barato. La variable *D_sinestud* no es significativa por lo que su análisis carece de validez. Cuando sopesamos la relación entre *ascensor* y el número de planta, ocurre que el signo de la estimación del parámetro que acompaña a *ascensor* es negativo (-52,02) y menor que el parámetro que acompaña a la ficticia multiplicativa (62,89). Es decir, cuando la vivienda no tiene ascensor, a mayor planta, menor precio del alquiler. Mientras que, si el edificio tiene ascensor, cuanto más alta se sitúe la vivienda, más caro es el precio del alquiler. Concretamente, vemos que cuando la vivienda no tiene ascensor, por cada planta que sube la localización de la vivienda, se reduce el precio del alquiler en 57,24€. Mientras que, si el edificio tiene ascensor, estar una planta más arriba aumenta el precio aproximadamente en 5,65€.

En segundo lugar, dadas unas determinadas características de la vivienda podemos predecir el precio que tendrá su alquiler.

Usando el modelo D, realizamos predicción extramuestral suponiendo los siguientes cinco casos:

1. La vivienda 1 se encuentra en Delicias, es decir, en el distrito 3 por lo que *D_población* va a ser igual a uno y *D_sinestud* va a ser igual a cero. Tiene ascensor, un sólo baño, 70 metros cuadrados y se encuentra en un quinto.
2. La vivienda 2 está en la zona del Centro, por lo que no cumple ninguna de las características planteadas (ambas igual a cero). Tiene ascensor, dos baños, 100 metros cuadrados y está ubicada en un cuarto.
3. La vivienda 3 está por la zona de Universidad, distrito 4. No tienen alta población ni un gran número de personas sin estudios. La vivienda tiene ascensor, un baño, 55 metros cuadrados y se encuentra en un quinto.
4. La vivienda 4 está en San José. Las variables *D_población* va a ser igual a uno y *D_sinestud* va a ser igual a cero. El piso no tiene ascensor y está en un tercero. Tiene 62 metros cuadrados y un baño.
5. La vivienda 5 se encuentra en el distrito 10, Actur-Arrabal. El piso no tiene ascensor, pero está en un primero. Tiene 80 metros cuadrados y dos baños.

Tabla 7.1. Predicción extramuestral con el modelo D.

	Vivienda 1	Vivienda 2	Vivienda 3	Vivienda 4	Vivienda 5
cte	320,012	320,012	320,012	320,012	320,012
renta_hogar	109,0947	165,2313	151,5852	109,0089	124,4061
metros	261,73	373,9	205,645	231,818	299,12
ascensor	-52,02	-52,02	-52,02	0	0
planta	-286,2	-228,96	-286,2	-171,72	-57,24
ascensor_planta	314,45	251,56	314,45	0	0
baños	0	102,24	0	0	102,24
Población	-38,93	0	0	-38,93	-38,93
D_sinestud	0	0	0	0	0
PRECIOS	628,1367	931,9633	653,4722	450,1889	749,6081

Por último, realizamos predicción extramuestral con el modelo F con los mismos casos que en la predicción anterior:

1. La vivienda 1 se encuentra en Delicias, es decir, en el distrito 3 por lo que $D_{densalta}$ va a ser igual a uno. Tiene ascensor, un sólo baño y 70 metros cuadrados.
2. La vivienda 2 está en la zona del Centro, por lo que $D_{densalta}$ va a ser igual a uno. Tiene ascensor, dos baños y 100 metros cuadrados.
3. La vivienda 3 está por la zona de Universidad, distrito 4. No tienen alta densidad de población. La vivienda tiene ascensor, un baño y 55 metros cuadrados.
4. La vivienda 4 está en San José, distrito 5. Las variables $D_{densalta}$ va a ser igual a uno. El piso no tiene ascensor, tiene 62 metros cuadrados y un baño.
5. La vivienda 5 se encuentra en el distrito 10, Actur-Arrabal. El piso no tiene ascensor y tiene 80 metros cuadrados y dos baños.

Tabla 7.2. Predicción extramuestral con el modelo F.

	Vivienda 1	Vivienda 2	Vivienda 3	Vivienda 4	Vivienda 5
cte	118,474	118,474	118,474	118,474	118,474
renta_persona	130,4082	209,9799	183,3624	134,0352	138,5163
metros	270,69	386,7	212,685	239,754	309,36
ascensor	83,34	83,34	83,34	0	0
baños	0	99,484	0	0	99,484
D_densalta	64,12	64,12	0	0	0
PRECIOS	667,0322	962,0979	597,8614	492,2632	665,8343

Ante las mismas características, los precios obtenidos son ligeramente similares para el caso de las viviendas 1, 2 y 4 con una diferencia de 30-40€. Las predicciones del modelo D presentan precios más bajos en las viviendas mencionadas. Mientras que para las viviendas 3 y 5, el modelo D predice valores más altos que el modelo F.

Conclusiones

El objetivo de este trabajo ha sido tratar de explicar el precio de la vivienda de alquiler en Zaragoza. Su utilidad principal es entender mejor las circunstancias concretas que pueden influir en él, además de poder aproximarse al precio de una determinada vivienda conociendo sus características específicas. El aumento continuado de los precios del alquiler tiene consecuencias graves en nuestra sociedad. Afecta especialmente a los jóvenes y a su bienestar. Por ello, se ha intentado identificar qué factores influyen en el precio y en qué medida mediante la metodología econométrica. Es tan preocupante que los precios se disparen como que se desplomen, ya que otros sectores se verían afectados. Contemplar los dos fenómenos es vital por parte de los agentes para la toma de decisiones y para plantear una posible solución que otorgue estabilidad.

En el trabajo se ha empleado la metodología en cuatro etapas. La primera etapa es la especificación que consiste en concretar la relación entre el precio y los factores explicativos. La relación se cuantifica en la etapa de estimación y se aplica una batería de contrastes de hipótesis y otras medidas escalares de bondad de ajuste con la finalidad de garantizar la fiabilidad de los modelos considerados adecuados. Por último, se utiliza el mejor modelo con objetivos tanto de predicción como de toma de decisiones por parte de los agentes implicados (propietarios, agencias, potenciales inquilinos, instituciones...). Esta etapa depende del objetivo del investigador. En nuestro caso, el principal objetivo es el análisis estructural pero también hemos realizado predicción.

En los modelos que finalmente superan la batería de contrastes podemos observar algunos fenómenos significativos. En primer lugar, las zonas con renta alta influyen positivamente sobre el precio del alquiler mientras que las zonas con renta baja hacen que los alquileres sean más baratos. Alquileres en barrios como el Casco Histórico, Delicias, San José, Las Fuentes o Torrero van a ser considerablemente más baratos que en otras zonas de Zaragoza. En segundo lugar, el efecto conjunto de renta y habitaciones muestra que un mayor número de habitaciones implica aumentos en el precio del alquiler de una vivienda que se encuentre en una zona con renta alta. Sin embargo, no se puede confirmar lo contrario. Tampoco se ha podido confirmar si influye de forma positiva que un piso estuviera amueblado en una zona rica. Suponíamos que, en una zona con renta alta, las viviendas podrían estar más decoradas o con muebles más nuevos. Descartamos la idea ya que ni *mueble* ni el producto entre *mueble* y *renta_hogar* eran significativas.

Los resultados que hemos obtenido con este estudio permiten resaltar factores que hacen que el precio del alquiler sea más o menos caro. Por ejemplo, si el objetivo es alquilar una vivienda al mínimo coste deberá presentar las siguientes características: no tener ascensor, ser pequeña (pocos metros cuadrados), tener sólo un baño, estar situada a más de un kilómetro del campus San Francisco y a más de cuatro kilómetros y medio del centro. En lo referente a las características de la zona, dicha vivienda tendrá que situarse en distritos con bajo nivel de renta, mucha población, baja densidad (un distrito grande, por ejemplo), alto porcentaje de población joven y de población sin estudios (que no superaron la escuela primaria).

No obstante, si la finalidad es comprar viviendas para después alquilarlas, el inversor querrá sacar el máximo beneficio, pero manteniendo precios atractivos. Dada la

significatividad que tiene la zona, convendría sacrificar otros aspectos como metros cuadrados o más de un baño y situar las viviendas en zonas con alto nivel de renta, poca población, mucha densidad, etc. De esta forma, los precios serían altos, pero no demasiado. Las viviendas medianas (en torno a 50 metros cuadrados) en zonas de estas características pueden resultar perfectas para una persona o para parejas jóvenes.

Resalta que cuanto más población haya en la zona, más barato será el alquiler de las viviendas. Sin embargo, a mayor densidad en la zona, más caros son los precios de los alquileres. Este efecto puede deberse a que la aglomeración permite tener mayor probabilidad de obtener productos y servicios, y esto derive en un mayor nivel de bienestar (Marín Verdú, 2019). La masificación de las ciudades favorece la actividad económica junto con otros factores que estimulan el crecimiento económico. Por tanto, los beneficios de una zona concentrada son superiores a los costes.

Los modelos que presentan mayor capacidad predictiva pueden aproximar el precio del alquiler de una vivienda dadas unas características determinadas. Por ejemplo, una vivienda situada en el barrio Delicias, en un cuarto con ascensor, una habitación y un baño, su precio de alquiler ronda los 600€ al mes. Cabe decir que es una aproximación con la información de la que disponemos.

Futuros trabajos podrían enfocarse en obtener datos más detallados, como otros gastos adicionales (comunidad, gas, luz, agua) o si los servicios básicos están dados de alta, la duración esperada del contrato, la antigüedad de la vivienda o reformas recientes, la contaminación acústica, etc.

Bibliografía

ANÓNIMO (2019): *Proyecciones macroeconómicas de la economía española (2019-2020): contribución del Banco de España al ejercicio conjunto de proyecciones del Eurosistema de diciembre de 2019*. Banco de España.

DÍEZ MUÑOZ, MIGUEL (2017): *¿Cómo afecta el turismo al precio del alquiler a residentes? Caso aplicado a la ciudad de Barcelona*. Facultat d'economia i empresa, Barcelona.

GARCÍA MONTALVO, JOSÉ (2019): *Retos del mercado del alquiler en España. Alquileres, Desigualdad, Shocks Externos*, 269.

LÓPEZ OLLER, JOFFRE (2019): *Observatorio de emancipación. Balance general primer semestre 2019*. Consejo de la Juventud de España.

MARÍN VERDÚ, PATRICIA (2019): *¿Sabes cuál es el efecto de las economías de aglomeración y de escala en la ciudad?* BBVA.

M. F. ZARAGOZA (2014): *Los zaragozanos, expertos testadores de nuevos productos*. Herald, Zaragoza

OFICINA TÉCNICA DE EBRÓPOLIS (2018): *Zaragoza en datos: Informe global sobre la ciudad y sus distritos*. Asociación para el Desarrollo Estratégico de Zaragoza y su entorno, Zaragoza.

PULIDO SAN ROMÁN, ANTONIO (1983): *Modelos econométricos*. Pirámide, Madrid.

SEGUÍ CORTÉS, DAVID (2017): *Estimación de un modelo de precios hedónicos para viviendas localizadas en el casco urbano de la ciudad de Altea (Alicante)*. Escuela politécnica superior, Alicante

SEGURA, SERGIO M. (2019): *Web Scraping for Data Analysis in Python: a case study*. <https://suresrm.github.io/2019/11/01/introduccion/>

TRÍVEZ BIELSA, FRANCISCO JAVIER (2004): *Introducción a la econometría*. Pirámide, Zaragoza.

VIAÑA, DANIEL (2019): *La desaceleración se cierne sobre España: el PIB crecerá ya menos de un 2% en 2020*. El Mundo, Madrid.

WOOLDRIDGE, JEFFREY M. (2013): *Introducción a la econometría. Un enfoque moderno*. Cengage Learning, México.